

# PadhAI: Variants of Gradient Descent

## One Fourth Labs

### Running mini-batch gradient descent

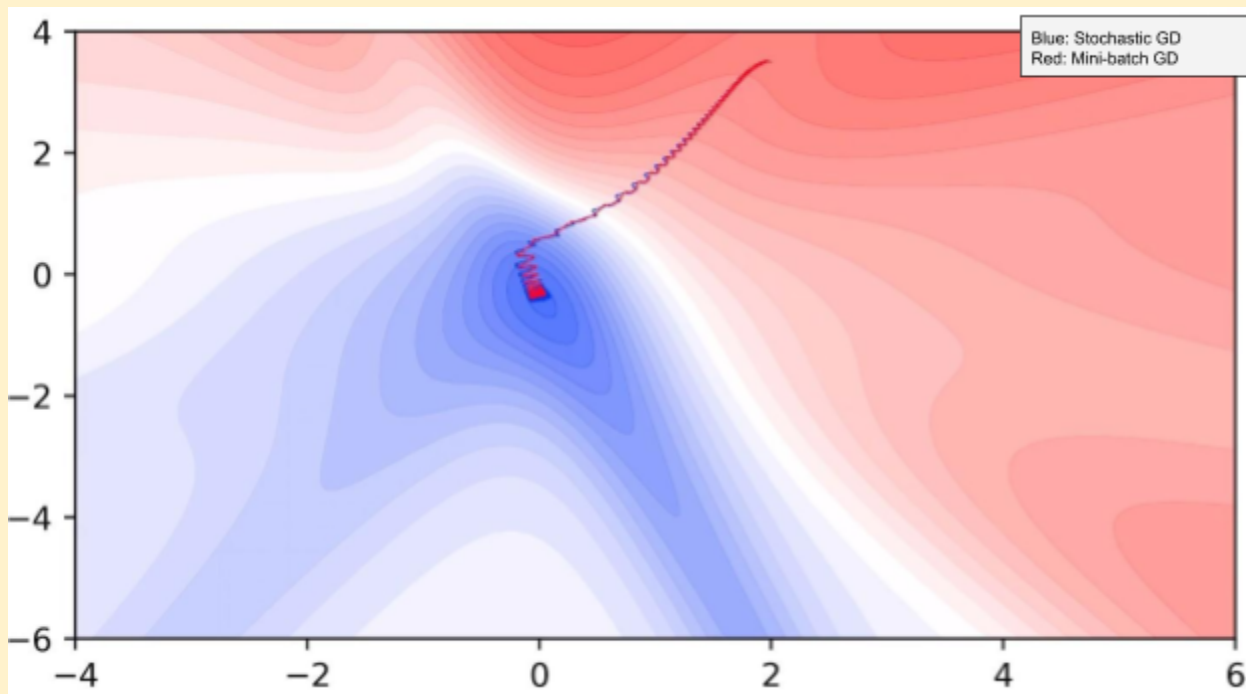
Doesn't it make sense to use more than one point or a mini-batch of points?

1. Let's look at the python implementation of mini-batch GD

```
def do_mini_batch_gradient_descent():
    w, b, eta, max_epochs = -2, -2, 1.0, 1000
    mini_batch_size = 10
    num_points_seen = 0
    for i in range(max_epochs):
        dw, db = 0, 0
        for x, y in zip(X, Y):
            dw += grad_w(w, b, x, y)
            db += grad_b(w, b, x, y)
            num_points_seen += 1

        if num_points_seen % mini_batch_size == 0:
            w = w - eta * dw
            b = b - eta * db
```

2. Now let us look at the 2D visualisation of mini-batch superimposed over stochastic GD



3. Here, we can observe that even though the plot oscillates for mini-batch GD, it is still considerably less than with stochastic GD, evidenced by the red plot lying entirely within the blue plot.
4. As we increase the batch size, the stability of the curve also improves, resulting in better estimates of the gradient
5. Recommended batch size is 32, 64, 128 etc.
6. The higher the batch size ( $k$ ), the more accurate the estimates.